DOCUMENT RESUME

ED 175 880 TM 008 327

AUTHOR Hanna, Gerald S.; Drahozal, Edward C.
TITLE Are Vocabulary Subscores Justified?
PUB DATE [77]
NOTE 6p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS -Context Clues; Educational Diagnosis; Elementary Secondary Education; Item Analysis; *Scores; Test Reliability; *Test Results; *Verbal Tests; *Vocabulary; Vocabulary Skills
IDENTIFIERS *Nelson Reading Test; *Subscores

ABSTRACT

Although vocabulary tests have not been well suited to subdivisions that yield scores for different attributes, they have often been used in this manner. The question of whether such subscores can be justified from a measurement standpoint was studied. Items of a vocabulary test were classifed into three categories: the words were presented in isolation, in phrase context, or in passage context. Subscores based on these three kinds of tasks were obtained from samples of approximately 3,000 students in each of grades 3 through 9. The subscores were then correlated. After correction for attenuation, the intercorrelations clustered sufficiently close to 1.00 to suggest that the separate subscores would not likely serve productive purposes. The two kinds of questions most frequently found in standardized vocabulary tests are isolated words and words in phrase context. The findings of this study indicated that these two item types offer alternative ways of measuring the same thing. It does not seem to matter whether a vocabulary test presents vocabulary words in isolation, in phrase context, or in a combination of the two. (Author/RD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gerald S. Hanna

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ARE VOCABULARY SUBSCORES JUSTIFIED?

Gerald S. Hanna
Kansas State University

Edward C. Drahozal
Houghton Mifflin Company

ED175880

With automated scoring services, it has become increasingly easy for publishers to offer numerous subscores as optional scoring services for various standardized tests. Many local school data processing operations likewise provide a large number of subscores to teachers and other school personnel. Although vocabulary tests have not seemed especially well suited to this kind of subdivision, they have not been immune to the trend. At least one major achievement battery offers for its vocabulary test an optional scoring service in which subscores and normative data are provided for items presented in phrase context and for questions cast in story context.

In developing the vocabulary section of a reading test, the present authors became interested in determining whether the items could be scored in such a manner as to yield subscores that measured different attributes. Wouldn't a teacher be interested in knowing that Jack could select synonyms of words presented in paragraph context at only the 25th percentile? Or that Jill's greatest vocabulary deficiency was in recognizing synonyms of words presented in phrase context? Wouldn't this kind of diagnostic information aid instruction? Wouldn't the availability of such scores also help to sell tests?

Indeed it would! But can a vocabulary test provide this kind of information? The research question addressed in this paper is whether such subscores are justifiable from a measurement standpoint.

Rationale

Correction of intercorrelations among scores for attenuation provides a simple way to study the extent to which two or more test scores measure different things and are therefore of potential use in profiling examinees' attributes. Obtained correlations are, of course, always lowered or attenuated by error of measurement of each of the variables upon which they are based. Correlation coefficients that have been corrected for attenuation can be viewed as estimates of the extent of the relationship between true scores. If correlation coefficients corrected for attenuation cluster around 1.00, then the respective scores measure the same thing (or at least can be considered interchangeable). To the extent that the corrected correlation coefficients are less than 1.00, then the scores measure different attributes. Imperfect intercorrelations after correction for attenuation are therefore a necessary (but not sufficient) condition for discriminant validity of all scores reported for a test.

For example if it were appropriate to report separate scores for reading comprehension of social studies material and reading comprehension of natural science content, then it would be essential that scores not be perfectly correlated after correction for attenuation. Likewise, if it is defensible to develop and use vocabulary subscores, it is vital that these subscores have corrected correlation coefficients substantially less than 1.00.

TM008 327

Procedures

Test Used. The Word Meaning Section of Form 3 of The Nelson Reading Skills Test (Hanna, Schell, & Schreiner, 1977b) was used in this research. The vocabulary items of this subtest are presented in three different ways: (1) in *isolation*, (2) in *phrase* context, and (3) in *passage* context. In the first case, students determine word meaning when an isolated word is used as the stem of a multiple-choice item. The student examines the word and locates a synonym for it from among the four choices. In the second case, the stimulus word is presented in a phrase context. The purpose of the phrase is either to require the examinee to determine from context a particular meaning of the word or to reduce the difficulty with which the more general or only meaning of the stimulus word is recognized. In the third case, word meaning is assessed within passages of a reading comprehension test. Students determine the meaning of a specific word in the passage context. One or more sentence meanings or thoughts may have to be comprehended in order to select the best option (Hanna, Schell, & Schreiner, 1977a).

Table 1 summarizes for each of the three levels of the test the number of Word Meaning questions of each category. The grades to which each level was administered are also shown.

Table 1

Number of Items by Category

Test Level	Grades	Isolation	Phrase	Passage
A	3 and 4	13	9	3
B	5 and 6	13	13	5
C	7, 8, and 9	15	14	7

Sample. Data from the test's entire spring standardization sample were used. This sample consisted of about three thousand students in each of grades 3 through 9. Sampling was organized to provide proportional representation of the national population of students in these grades with regard to socio-economic status of community, enrollment of school district, and geographical region of the country. More detailed information about the sample is presented elsewhere (Hanna, et al., 1977a).

Scoring. For the purpose of this study, separate number-right subscores for each examinee were obtained for (1) words in isolation, (2) words in phrase context, and (3) words in passage context.

Estimating Reliability. For each of these three subscores at each grade, the Kuder-Richardson 20 reliability coefficient was computed.

Establishing Attribute Independence. At each grade level, the intercorrelations among the three kinds of subscores were computed. Kuder-Richardson formula 20 reliability estimates were then used to correct these obtained intercorrelations for attenuation.

Results and Discussion

Table 2 reports the findings for each grade. The data above the diagonal of each part of the table are the obtained correlations among the three subscores. Scores of the words in isolation and scores of words in phrase context correlate higher with each other than either does with scores of words in passage context; this is partly an artifact of the number of items included in each subscore. The diagonal of each section of Table 2 contains in parentheses the K-R 20 reliability estimates of these three subscores. Here too the lower reliability estimates of the scores for words in passage context is largely an artifact of the relatively small number of items in this scale. The section below the diagonal contains the intercorrelations corrected for attenuation. These corrected values eliminate the systematic influence of the subscore length artifact.

The size of the corrected correlation coefficients between scores of words in isolation and words in phrase context range from .98 to 1.04. The median is 1.02. It is obvious that these seven corrected intercorrelations are essentially perfect. Deviations from unity appear to reflect nothing more than minor sampling errors; when the correlation between two scores is overestimated in a given sample and/or when the reliability of one or both of the scores is underestimated in the sample, then the correlation corrected for attenuation--the *estimated* correlation between true scores--can exceed 1.00.

The corrected correlation coefficients between scores for words in isolation and words in passage context range from .86 to 1.00 and have a median of .94. The corresponding corrected correlation coefficients for scores of words in phrase context and words in passage context range from .84 to .94 and have a median of .92. What was measured in these samples by words in passage context is highly related, but not identical, to what was measured by the other two item types.

Educational Significance

The two kinds of questions most frequently found in standardized vocabulary tests are isolated words and words in phrase context. The findings of this study suggest that these two item types offer alternative means of measuring the same thing. It does not seem to matter whether a test presents vocabulary words in isolation, in phrase context, or in a combination of the two.

Table 2

Obtained Correlations, Reliability Estimates, and Corrected Correlations

Grade 3
N = 2984

	Isolation	Phrase	Passage
Isolation	(.81)	.81	.58
Phrase	1.04	(.76)	.56
Passage	.93	.92	(.48)

Grade 4
N = 3310

	Isolation	Phrase	Passage
Isolation	(.79)	.80	.61
Phrase	1.02	(.78)	.59
Passage	.86	.84	(.62)

Grade 5
N = 2958

	Isolation	Phrase	Passage
Isolation	(.83)	.83	.69
Phrase	1.02	(.82)	.68
Passage	.94	.94	(.65)

Grade 6
N = 3301

	Isolation	Phrase	Passage
Isolation	(.82)	.83	.68
Phrase	1.02	(.82)	.69
Passage	.91	.92	(.68)

Grade 7
N = 3414

	Isolation	Phrase	Passage
Isolation	(.81)	.79	.69
Phrase	.98	(.81)	.65
Passage	.98	.92	(.62)

Grade 8
N = 3135

	Isolation	Phrase	Passage
Isolation	(.80)	.80	.68
Phrase	.98	(.83)	.66
Passage	1.00	.94	(.59)

Grade 9
N = 2314

	Isolation	Phrase	Passage
Isolation	(.81)	.82	.69
Pharse	.99	(.84)	.66
Passage	.97	.90	(.63)

On the other hand, scores for words in passage context seem to measure something slightly different than scores of either of the other item types studied. The median correlation of scores of words presented in passage context with scores of isolated words was .94 and with scores of words in phrase context was .92. Clearly most (88 percent and 85 percent respectively) of the reliable variance is common variance. If one can generalize from the test items used in this study, it seems very doubtful that the slight amount of nonidentical reliable variance in grades 3 through 9 would justify reporting subscores for words in passage context and subscores for the other kinds of vocabulary items studied. With subtests of great length, it is possible that differences for individual examinees could be detected with modest reliability. But the time spent in obtaining such reliable subscores would seem to be ill-advised in view of the limited reliability of the resulting information and the limited educational use to which such information would likely be put.

Test and curriculum specialists in school districts are often pressured to provide item and subscore information of the type described in this paper for survey tests. Although modern scanning and data processing equipment can provide such data at a modest cost, the monies so spent do not yield commensurate educational benefits to teachers and other school personnel.

References

Hanna, G. S., Schell, L. M., & Schreiner, R. Teacher's Manual: The Nelson Reading Skills Test. Boston: Houghton Mifflin Co., 1977a.

Hanna, G. S., Schell, L. M., & Schreiner, R. The Nelson Reading Skills Test. Boston: Houghton Mifflin Co., 1977b.